A RELATIVE STUDY OF DIFFERENT MACHINE LEARNING CLASSIFICATION ALGORITHMS TO FORECAST THE HEART DISEASE

BUSHRA MEMON[1], SUMBUL GHULAMANI[1*]

[1]Dept. Of Computer Science, SZABIST Hyderabad Campus, Pakistan

**Corresponding author: sumbul.ghulamani@hyd.szabist.edu.pk*

ABSTRACT: Cardiovascular disease is comprehended as heart disease, and it covers different cases that impact the heart and has been the direct cause of death. It allies numerous risk elements in heart disease and the necessity to build practical strategies for earlier diagnosis to manage the condition promptly. The machine learning classifier has evolved significantly in the medical database, particularly for diagnosing disease. Nowadays, numerous organizations utilize these machine learning strategies to improve medical diagnostics for the earlier prognosis of conditions. This paper summarizes these machine learning algorithms in disease prediction and computation. Classification algorithms considered are logic regression, Naive Bayes, K-nearest neighbor (K-NN), Support vector machine (SVM), Decision Tree (DT), K-means clustering, and Random forest (RF). This work reviewed 20 papers from 2018 - 2021 that employed Classifier to detect specifically heart diseases in the medical sector during the last four years.

KEY WORDS: *Cardiovascular Disease, detection, classifiers*

1. INTRODUCTION

Disease diagnosis is determining which illness describes an individual's symptoms. The situation is most challenging in the diagnosis because some signs and indications are not clear. The essential fact is identifying ailments for treating a disease (Qaiser et al., 2015). In the human body, heart is an integral part. A muscular organ pumps blood within the body and holds its cardiovascular system (Yahaya et al., 2020). Cardiovascular comprises blood plates, including veins, capillaries, and arteries, which build a formidable blood circulation network throughout the human body (Yahaya et al., 2020).

Atherosclerosis and hypertension stand among the most common causes of heart disease. Atherosclerosis is an ailment that happens in the artery walls if plaque accumulates. Moreover, that build causes narrow arteries and composed the more blood flow. Blood clot forms (Joshi & Nair, 2015). Nowadays, cardiovascular diseases (CVD) are a hot subject in the Healthcare industry globally (Qaiser et al., 2015). World Health Organization (WHO) proclaimed the newest data (Kalamma & Shreedhara, 2016). According to the heart disease reports,

progression of diseases that impacts the heart, today, CVDs are the primary element of death globally, with 17.9 million deaths yearly (Bharti et al., 2021). Diseases and health concerns like cardiovascular and heart syndrome significantly impact one's health and, if neglected, the direct causes of death (Aswale & Shetgaonkar SRIEIT-Goa Shailendra Aswale, 2021). Thus, medical scientists are interested in advanced model techniques for detecting heart ailment based on machine learning classifiers (Ibrahim & Abdulazeez, 2021). Physicians require accurate prediction results in their patient data during the disease diagnoses (Alanazi et al., 2017). There is a need to use machine learning for medical practitioners to predict heart diseases before they occur in their patients. Numerous research existed (Bharti et al., 2021). The models of various classifiers of machine learning were utilized to forecast heart diseases (Ao et al., 2014).

Disease analysis determines which disease describes a person's symptoms, and this is the most challenging situation to diagnose heart diseases. Categories of epidemics raise various important features concerning the cure of infections. Machine learning classifiers prediction based on earlier training data point out the specific area to help detect the heart ailments (Bharti et al., 2021; Borges et al., 2021).

The scope of predictions in machine learning is implied by identifying trends in the patient data to enhance their healthcare. Today, Artificial intelligence and machine learning have been used in the medical sector (Borges et al., 2021). In developing countries like South Africa, Cardiovascular Disease (CVD) has become a leading death problem. CVD is the only highest killer globally considering all (Damen et al., 2016; Kalamma & Shreedhara, 2016; Research & 2018, 2018). Using machine learning opportunities, an effort to utilize patients' information, experience, and clinical screening to diagnose heart attacks was treasured for advanced earlier prediction (Ao et al., 2014). These approaches provide the advantages, grouping the patients holding a similar kind of health disease. They can provide effective treatments, monitor the availability of pharmaceutical solutions over patients at more inexpensive cost, offer reliable healthcare strategy, diminishing the time towards preventive medication and detect causes of diseases, best to identify the methods and experts for making an efficient healthcare policy (Qaiser et al., 2015). Machine learning is broadly used in the patient data record to find the best disease prediction on large medical records. This advanced approach can be more beneficial when diseases data is properly analyzed (Dwivedi et al., 2021; Smiti, 2020).

It is universally recognized that machine learning and artificial intelligence play essential roles in the medical industry. Multiple approaches of machine learning, and deep learning simulations are being used, identifying the disease and analyzing or forecasting the outcome (Bharti et al., 2021). The machine learning technique provides a way to collect valuable data and details from large databases (Collins et al., 2021). Furthermore, the data mining strategies that mainly utilize such classifiers are regression, k-clustering, and analysis techniques such as Naive Bayes, Decision tree, Random Forest, and K-NN (nearest neighbor) that are applied to detect different heart disease attributes (Kalamma & Shreedhara, 2016).

This research reviews the machine learning classification techniques employed to predict healthcare data related to cardiovascular. The various classifiers are Decision Tree, Regression, Naive Bayes, K-NN (nearest neighbor) classifier, Random Forest, Support Vector Machine, k- means clustering. This paper explores several past studies considering research papers from 2018 to 2021, which only focused on heart disease. Machine learning classifiers are utilized to determine multiple statuses in the medical field in the last four years. They have developed the best prediction method for accuracy on specific dataset to improve the early prediction of pharmaceutical disease diagnostics.

2. MACHINE LEARNING (ML) KINDS

ML stands for artificial intelligence, which permits the machine to consider as the human thinks and construct determination without any interaction of human beings (Ibrahim & Abdulazeez, 2021). Its focus is to develop techniques and make predictions based on experience. Its feature extraction method trains the scheme to get facts from the algorithms to construct structures using the different datasets. The intake of new data predicts heart ailment and finds the missing values in the dataset. It assembles valid predictions for new datasets. Numerous kinds of ML approaches are classified (Banu et al., 2019; Dwivedi et al., 2021), as shown in Fig.1.

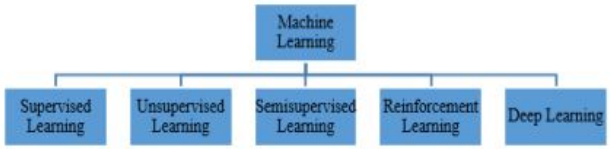

Fig. 1. The machine learning types (Shah et al., 2020)

This section briefly describes kinds of ML. In the supervised learning approach, the datasets are prepared on labeled data. This signifies that the algorithm predicts data from input data that generate the required output. In the unsupervised learning, classifier containing data input understands the innate system through input (Shah et al., 2020). At the same time, semi-supervised classifier relies upon techniques (supervised and unsupervised). Utilizing both implies that some data are labeled and unlabeled. The algorithm has not utilized labeled data (Ibrahim & Abdulazeez, 2021). Deep Learning approach is a subcategory of ML, with multiple coatings. Each coating obtains data inputs from the layer, as mentioned earlier, and shows some required outcomes in subsequent layers (Kumar et al., 2020).

3. MACHINE LEARNING CLASSIFIERS

The different ML methods are applied to determine the earlier detection of heart diseases diagnoses. Classifiers founded in this study are written below: K-mean clustering, Deep Learning (DL), Decision Tree (DT), Logistic Regression (LR), Naive Bayes (NB), and K-NN (nearest neighbor), and Random Forest (RF).

3.1. Random Forest classifier:

This method is used to improve prediction power, techniques that integrate many Decision Tree methods and combine the random selection of features for each of the trees from our dataset as a subset together and take the majority division from the trees and decide the classification on that. Furthermore, the power of numbers can help eliminate specific errors and delays in algorithm and construct better distinct learning algorithms. The major pros are that it can proficiently address massive quantity of data (Joshi & Nair, 2015). The parameter with the best prediction accuracy as the number of estimator increases is shown in Fig. 2.

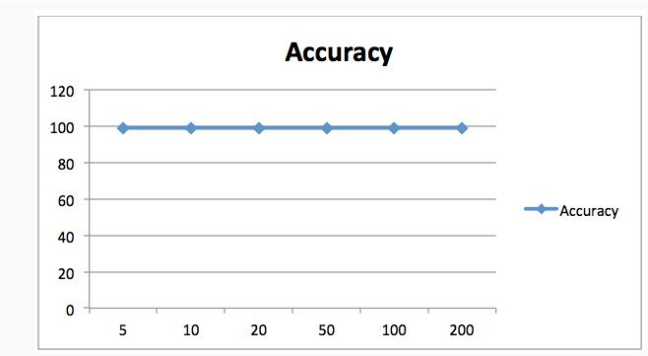

Fig. 2 . Random Forests classifier (*Random Forest Algorithm with Python and Scikit-Learn*, n.d.)

3.2. K-nearest neighbors (KNN) Classifier:

The instance-based approach stands as the k-nearest neighbor, is a classification procedure that recognizes models. The technique discovers the new instance in points mentioned earlier as a model that fits the unknown sample. For the forecast, the present model considerably matches the previously examined instance. The K-NN sorts the cases via the K nearest neighbors. The training rate of this K-NN classifier, as shown in Fig.3., was observed as faster but relatively slow whenever they receive the dataset and was significantly evaluated for all instances (Bharti et al., 2021).

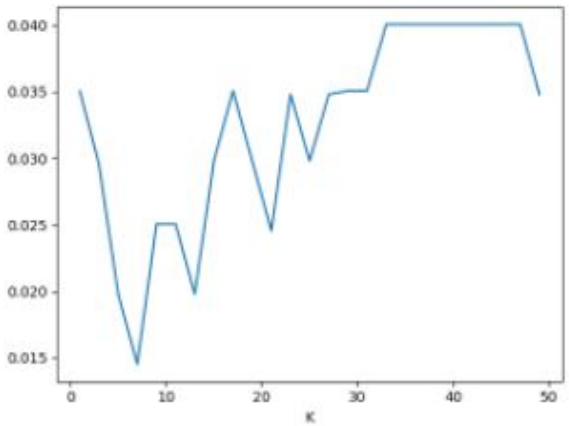

Fig. 3. k-neighbor classifier (*K-Nearest Neighbor(K-NN) | Python - AI ASPIRANT*, n.d.)

3.3. Decision Tree classifier:

A Decision tree is a technique that contains source nodes, branches, and shoot nodes. Respectively, inner node represents an attribute, nodes show the trial outcomes, and per leaf node holds the class title. They picked attributes and set them as root first. For every attribute, the value breaks the datasets into subsets. The recursive process repeats that per division, employing only those who reach the team. Whenever all models contain the exact classification, the development of the tree can stop (Bharti et al., 2021; Kumar et al., 2020). Fig. 4 shows some maximum depth data values to the representation of how the algorithm predicts the data points.

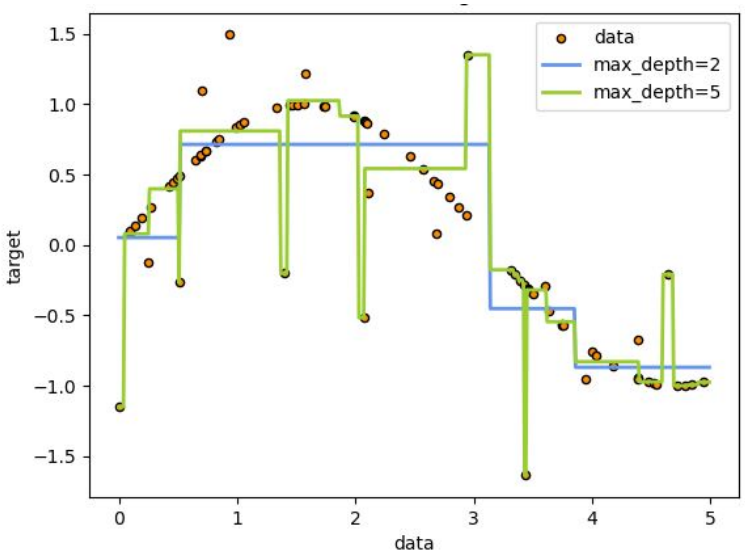

Fig. 4. Decision classifier (*Decision Tree Regression — Scikit-Learn 1.0.2 Documentation*, n.d.)

3.4. Logistic Regression (LR) classifier:

The LR stands as supervised learning utilizes this approach for translating a problem into the binary category. A mathematical approach measures and predicts provided data objects using the regression model. This classifier contains numerous essential attributes in terms of effective performance and computation (Ao et al., 2014; Bharti et al., 2021; Research & 2018, 2018). For the input elements, number of scaling instructed is as shown in Fig. 5.

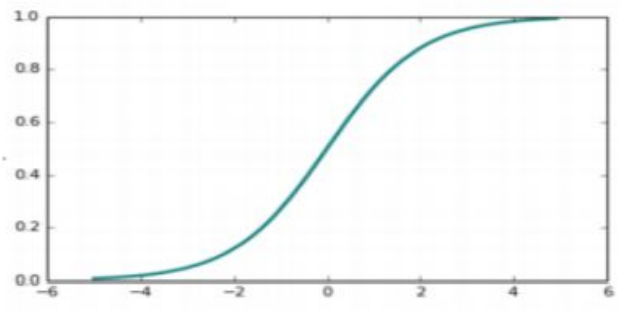

Fig. 5. Logistic function (Joshi & Nair, 2015)

3.5. K-Means Clustering classifier:

The k-means, an unsupervised learning methodology, is used is to clustering the nearest neighbor data based on resemblance. K technique is integer, numeral values known for the classifier to perform the clustering. The algorithm detects the right clustering points in the new data according to distance. The selection was performed randomly of this classifier centroids. Then, all nearest attributes were distributed to their nearest centroids created class. This technique repeats until you get the essential facts from the datasets (Ibrahim & Abdulazeez, 2021), as shown in fig. 6.

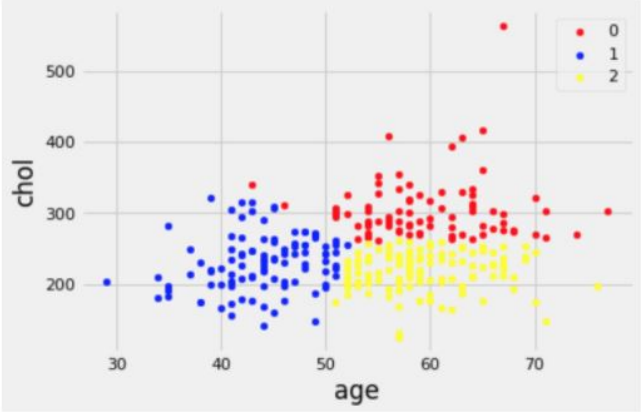

Fig. 6. k-clustering classifier (*K-Means Clustering for Analysis of Heart Disease | by Michelle Hottinger | Medium*, n.d.)

3.6. Deep learning(DL) classifier:

This algorithm permits the data learning into multiple classes of representation—the data in the model. The unknown factors in the intake allocation extract features during the training process. Classifiers discover the functional patterns of the data. Considerably, machines are trained to use the ML algorithms to construct the model. Deep learning techniques generate outcomes faster than

standard machine learning approaches by extracting complicated abstractions as data models through a hierarchical learning approach (Joshi & Nair, 2015).

3.7. Naive Bayes (NB) classifier:

The execution of this machine learning algorithm makes predictions in real-time quick, brief, and space efficient. This classifier of machine learning is a probabilistic algorithm and is implemented to make projections in real-time swift, fast, and space efficient. More minor training data can be utilized in discrete and continuous data in the Naive Bayes classification categories (Bharti et al., 2021; Shah et al., 2020), as shown in Fig. 7.

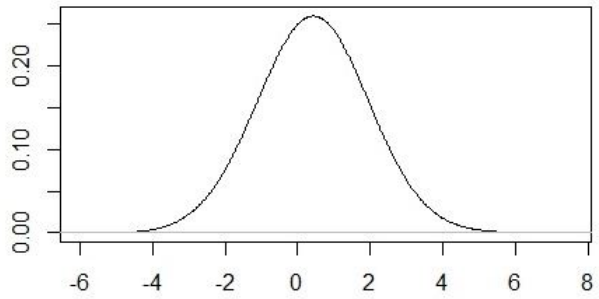

Fig. 7: Naive Bayes classifier (Bharti et al., 2021)

4. LITERATURE REVIEW

Many studies and work have been done on disease diagnoses using machine learning and data mining techniques (Qaiser et al., 2015; Aswale & Shetgaonkar SRIEIT-Goa Shailendra Aswale, 2021). The data mining methods are more helpful at predicting heart diseases. HD datasets of patients from the University of California, Irvine hold 14 attributes with 668 records. Experiments were conducted with three AI-based methods: Decision Tree, Naive Bayes, and Neural Network for determining heart disease. The experiment showed that the chest pain and cholesterol attributes reduced the efficiency of a decision tree for heart disease prediction. The decision tree analysis result was more accurate with 98.54% as compared to other AI approaches where Naive Bayes was 85.01% and Neural network was 81.83%.

In 2021, Bharti et al. (2021) targetted the field, pointing to the appearance of heart disease in the patient. ML classifiers of linear model selection used Logistic Regression, Random Forest, Decision Tree, and KN neighbors. Also combined with the deep learning approach in two ways using a sequential method and one more is a deep functional procedure; after comparing both approaches, the machine learning discovered the better accuracy in terms of effectiveness and efficiency of algorithms. This experiment utilized the confusion matrices on the UCI dataset and combined them with technology. Employing the deep learning approach, they found an accuracy of 94.2%. In 2021, the predicting machine technique was developed based on the Hybrid approaches which merge the decision tree and random forest classifiers (Kavitha et al., 2021). Dataset utilized Cleveland's heart for best accuracy to diagnoses, the heart attacks ratio.

In 2021, Jindal et al. (2021) presented machine learning classifier method that predicts the individual heart attack ratio. Patient diagnoses the disease or not based on prediction accuracy rate. The dataset from UCI was taken, and regression and k-nearest neighbor classifier were applied for prognosis to classify the aliment. In 2021, Pandita et al. (2021) diagnosed the patient records using a data mining approach to predict the best accuracy of heart disease. Dataset utilized for analyses was taken from Kaggle. They trained the data by employing the ML classifier for prediction and the K-nearest neighbor, Decision Tree, and Random Forest. In 2021, Sajja (2021) presented the framework for physicians to diagnose heart diseases efficiently. Model created a computer aided diagnosis technique. Data mining classifiers were applied to Cleveland dataset for prediction.

In 2020, Kumar et al. (2020) proposed a model to predicate (CVD) the machine learning ML classifiers like the Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), Support Vector machine (SVM), and K-nearest neighbors (K-NN). The Random Forest classifier's outcome had the most satisfactory precision in disease prediction than the other algorithms. In 2020, Gazeloğlu (2020) found that CVDs are among the most apparent diseases in human beings. They used eighteen machine learning processes and separated them and categorized into six classifications and three different feature selections by the experimental approach. After that, they analyzed the proposed methods through Weka, Python, and MATLAB program. Participants belong to 303 patients. All are composed of 14 variables. Analysis results showed that SVM without feature selection, and the ratio was found to be 85.148%, and the Correlation-based feature selection, Naive Bayes, and Fuzzy Rough Set with a balance of 84.818%. Also, they experimented using Chi-Square feature selection, RBF Network algorithm with an 81.188% ratio. All the methods were found to be the most successful algorithm.

In 2020, researchers compared the classifiers that have been employed, such as the Naive Bayes, Artificial Neural classifier, and Decision Tree to predict multiple heart disease datasets from geographical areas to analyze the accuracy of these algorithms (Yahaya et al., 2020). This analysis has 303 samples with 14 features. They have observed comprehensive research. The studies used the Cleveland heart disease dataset, minimizing morbidity and mortality rates due to CVDs. Many studies were performed with machine learning models to classify the heart disease of the diagnosis. In 2020, various machine learning approaches were implemented (G, 2020). The Python language was utilized for implementation. UCI dataset was taken for heart disease. After analyzing the algorithms such as Logistic Regression, Naive Bayes, Support Vector, Random Forest, and Decision Tree, the results showed that Random Forest gave more precision than the other five classifiers. It was identified that between 50 to 60 age group people suffer more from heart diseases, and a higher ratio was found in men between 80 to 100 years of age compared to women and between 60 to 80 age group of women are suffering from heart disease. The observed outcome is that the Random Forest classifier was found with the maximum precision of 91.8%.

In 2020, Shah et al. (2020) studied the data mining algorithms that predict the patient's heart attacks ratio and compared the best prediction method. Supervised

learning algorithms and data mining classifiers were applied to 20 datasets from medical practitioners to diagnose the heart attacks ratio in the patient. Algorithms showing the best outcomes in this model were found to be highest in K-nearest neighbors, Naive Bayes, and Random Forest prediction with accuracy of 99%. In 2019, Atallah et al. (2019) presented a predictive prototype for the disease of Heart-constructed on the achievable fact. The objective of the ML approach is to centralize a more minor in cost medical investigation accomplished in any provincial clinic sector. Correspondingly, to get benefit from this advanced technique by dedicating the doctors equipping the real-time data of patients. In 2019, Alotaibi (2019) conducted a comparative analysis to predict heart disease using some machine learning classifiers techniques. Cross-validation strategies were applied to the UCI Cleveland heart disease dataset. From the outcomes of the experiments, the most raised in the heart disease forecast was observed.

In 2019, Anitha et al. (2019) used a framework utilizing the supervised machine learning approach in the programming R platform. The dataset was taken from the University of California, Irvine, based on the 76 components and 303 selections. The analysis of the selected strategies testing outcomes demonstrated that the Naive Bayes classifier performed the best prediction to obtain the heart disease ratio. In 2019, Basha et al. (2019) suggested an analysis model to detect heart syndrome. These diseases forecast has some essential parameters that matter the most such as gender and age, the blood pressure, diabetes ratio of human health, and heart momentum. Data scientists utilized machine learning algorithms to efficiently diagnose heart syndrome. In safeguarding patients' lives, the researchers accomplished predictive work on enormous data to diagnose heart syndrome early. In 2019, Padmanabhan et al. (2019) created an Auto model by utilizing machine learning to predict the risk of cardiovascular diseases. Some essential points were analyzed regarding the total amount and time needed. The classifiers utilized were LR, SVM, RF, Boosting, NN, and KNN. Experiment on two different datasets, from UCI took 432 hours and 360 hours for CVDs.

In 2018, Hariharan et al. (2018) carried out comparative analysis on heart disease prediction by utilizing machine learning algorithms. Data were taken from USI, VA Long Beach comprising two hundred seventy samples and 12 features. A confusion matrix assessed the model based on accuracy, understanding, and particularity. Experimental outcomes indicated that Support Vector Machine is more accurate than other ML algorithms in classifying heart ailment patients. In 2016, Damen et al. (2016) highlighted an integrated framework utilizing various algorithms to predict heart diseases. The UCI dataset utilized was taken from Long Beach. Experiment outcomes of the Naive Bayes were 79.12% valid whereas escalated, 87.91% when combined to PSO classifier. In 2018, David et al. (2018) employed three data mining techniques to produce a prediction system to analyze the possible rates of cardiovascular diseases. This paper intends to determine the best prediction model with the highest accuracy when expected, and individuals classify unexpected facts. The UCI dataset from Long beach was utilized for models' training and testing processes. The dataset was 80% and 20% for training and testing models to get better prediction outcomes.

In 2015, Qaiser et al. (2015) evaluated the performance of the algorithms in healthcare for heart diseases. Datasets were from the University of California, Irvine. The patient dataset contains 303 records from the Cleveland Clinic Foundation. The Hungarian dataset containing 294 patient records was combined to analyze the accuracy of algorithms. The study used experimental approach in two different scenarios using data mining algorithms. The first scenario used all attributes, and the second scenario used the features. The experiments were conducted using WEKA. The algorithms incorporate different accuracy metrics. These experiments showed that the K-nearest neighbor and Naive Bayes classifiers have the most elevated accuracy and the fastest execution time compared to other algorithms.

5. DISCUSSION

This paper describes all the studies performed by the researchers to specify heart disease that utilized the machine learning classifiers. Table 1 summarizes several past papers in this comparative study, assessing research papers from 2018 to 2021, which only focused on heart disease. Classification algorithms considered are Logic Regression, Naive Bayes, K-NN (nearest neighbor), Support Vector Machine (SVM), Decision Tree, K-means clustering, and Random Forest (RF). Year, patient record (datasets), and percentage of classifiers, predicting accuracy, outcomes, and findings on heart diseases are highlighted.

6. CONCLUSION

The extensive literature reviewed for this research summarizes the various kind of machine learning classifiers for predicting cardiovascular disease. Most analyses have employed the Cleveland heart disease dataset. Few studies utilized other data sources with finite heart disease features in different conditions like heart failure, cardiovascular, and heart syndrome. A list of outcomes found are tabulated.

Forecast specifically heart diseases used different machine learning classifiers. After comparing 20 papers modeling the forecasted diseases, many classifiers were found to have the good predicting accuracy, but K-nearest neighbors, Random Forest, and Decision Tree have more satisfactory accuracy than other classifiers.

Table 1: Summary of the papers work performed by the experimenters reviewed in this study

Ref	year	Disease	Machine Learning algorithm	Dataset	Result	Findings

(Aswale & Shetgaonkar SRIEIT-Goa Shailendra Aswale, 2021)	2021	Heart attack	LR Logistic Regression classifier, RF Random Forest, DT Decision Tree, and K-NN K-nearest neighbor.	UCI machine learning repository data.	Accuracy level of Decision Tree=98.54%, Neural Network= 81.83%., Naive Bayes=85.01% and Random forest=99%	Interpreted heart attacks ratio in the patient, the random forest classifier achieved the highest prediction accuracy level.
(Bharti et al., 2021)	2021	Heart	DT Decision Tree, LR Logistic Regression, RF Random Forest, and K-Nearest neighbor.	UCI machine learning repository patient data.	Accuracy level of Random Forest classifier =94.2%.	Efficiently diagnosed the heart disease.
(Biology & 2021, 2021)	2021	Cardiovascular	Naive Bayes and LR, Sequential minimal optimization (SMO), AdaBooostM1 and Data science, Logistic regression, bagging and REP Tree, JRip and Random Forest	UCI Cleveland heart dataset	Accuracy level of SMO= 85.148, Logistic regression= 84.818 and AdaBoostM1=0.90, Naive Bayes = 0.184.	Predict the risk of cardiovascular diseases.
(Kavitha et al., 2021)	2021	heart attacks and coronary artery diseases	Hybrid based on Decision tree and random forest	UCI machine learning repository	Accuracy level=88.7%	The predicting machine technique is developed based on the Hybrid approaches with marge the Decision tree and random forest classifiers.
(Jindal et al., 2021)	2021	Heart attack	applied regression and k-nearest neighbor classifier	UCI patient data	Accuracy level =87.5%	predicts the individual heart attack ratio
(Pandita et al., 2021)	2021	Heart disease	K-NN k-nearest neighbor, DT decision tree, and RF random forest	Kaggle data	100% accuracy received	Diagnoses the patient records using a data mining approach to predict the best accuracy of heart disease.
(Sajja, 2021)	2021	Heart diseases	SVM Support vector machine, RF Random Forest,	Cleveland data	Support vector classifier	Physicians to diagnose heart diseases efficiently.

			and K-NN k-nearest neighbor.		accuracy is most achievable.	
(Almustafa, 2020)	2020	Heart disease	SVM, k-Nearest Neighbor, NB Naive Bayes, DT Decision tree J48, and J Rip, Adaboost, Stochastic Gradient Descent and Decision Table	Patient data Cleveland, Hungary, Switzerland, and Long Beach	Accuracy level of K-Nearest Neighbor = 99, Decision tree J48 =98.0, Rip=97.2 and decision Table= 93.8	Efficiently forecast the heart disease ratio in patients.
(Yahaya et al., 2020)	2020	Cleveland heart disease	Naive Bayes, Decision Tree, and Artificial Neural Network	Cleveland heart disease dataset	Comparatively, examine accuracy of the classifiers.	The approach for CVDS is to minimize the rates.
(Gazeloğlu, 2020)	2020	Cardiovascular	Support vector machine, Naive Bayes NB, Fuzzy Rough Set, RBF Network classifier, Chi-Square Correlation-based and Feature Selection.	Kaggle, Data	Accuracy level of SVM =85.1% Naive Bayes, and Fuzzy Rough Set =84.8%. RBF Network algorithm= 81.1%	The recommended method helps specialists utilized RBF in hospitals to classify heart disease.
(Kumar et al., 2020)	2020	Cardiovascular	(DT), (LR), (SVM), (RF) and(K-NN)	The (NIDDK) National institute of diabetes and digestive and kidney disease.	Accuracy Level of KNN=68.5%, DT =74. %, LR=74 SVM=77 and RF=88%	Analysis of the classifiers surpasses the classifying CVD patients.
(G, 2020)	2020	Cardiovascular	(Naive Bayes classifier) NB, (logistic regression) LR, (random forest) RF, (Support vector machine) SVM, (decision tree) DT, and K-NN (Nearest Neighbor).	UCI machine learning repository data.	Accuracy level of random forest classifier among all=91.8%.	The decision support designs to forecast CVD in patients.

(Atallah et al., 2019)	2019	Heart	LR, (RF), (K-NN) and Ensemble approach.	UCI heart data.	Accuracy level of Ensemble =90%, Random Forest= 87%, KNN=87% and Logistic Regression=87%	The ensemble representative acquired heightened accurateness after analyzing each classifier.
(Padmanabhan et al., 2019)	2019	Cardiovascular	LR, SVM, RF, Boosting, N-N, and K-NN.	UCI machine learning repository CVD's data.	UCI takes 432 hours and 360 hours for CVD's.	Auto model by utilizing machine learning to predict the risk of cardiovascular diseases.
(Basha et al., 2019)	2019	Heart Syndrome	Random forest RF, DT, K-NN, NB, and SVM Classifier	Kaggle Patient data.	Accuracy level of K-NN=85%, DT=82%, SVM=82%, RF=81% and NB=80%	Predictive work on enormous data for earlier analysis of heart syndrome.
(… & 2019, 2019)	2019	Heart failure.	RF, NB DT, LR, and Rapid Miners.	UCI Cleveland heart disease data.	Accuracy level of Support vector machine=93%, Decision Tree=92.3%.	The most raised in the heart failure forecast disease.
(Damen et al., 2016)	2019	Heart	SVM, K-NN, and Naive Bayes NB	University of California, Irvine, data	Accuracy level of Naive Bayes = 86% more accurate than the Support vector machine and KNN.	The suggested framework to signify heart ailment in patients.
(David et al., 2018)	2018	Cardiovascular	Random Forest, Naive Bayes, and Decision Tree.	UCI dataset from Long beach	The Random Forest classifier conducted better accuracy than the Naive Bayes and Decision Tree to forecast heart disease	Data mining techniques to produce a prediction system to analyze the possible rates of cardiovascular diseases.
(Research & 2018, 2018)	2018	Heart	Naive Bayes (NB)and Particle Swarm Optimization (PSO) classifier	UCI data has been taken from Long Beach	Accuracy level of Naive Bayes = 79% and PSO=87%	An integrated framework utilized algorithms to predict heart diseases

(Harihara n et al., 2018)	2018	CVD'S	SVM, DT, and k-nn.	USI, VA Long Beach	Accuracy level of Support Vector Machine= 92.1%, and K-Nearest Neighbor= 83.2%.	Comparative analysis on heart disease prediction.

REFERENCES

Alotaibi, F. S. (2019). Implementation of machine learning model to predict heart failure disease. *International Journal of Advanced Computer Science and Applications*, *10*(6). https://pdfs.semanticscholar.org/a74f/d8c51251e8c6126a1527e545bd78860a10f9.pdf

Alanazi, H. O., Abdullah, A. H., & Qureshi, K. N. (2017). A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care. *Journal of Medical Systems 2017 41:4*, *41*(4), 1–10. https://doi.org/10.1007/S10916-017-0715-6

Almustafa, K. M. (2020). Prediction of heart disease and classifiers' sensitivity analysis. *BMC Bioinformatics*, *21*(1). https://doi.org/10.1186/s12859-020-03626-y

Anitha, S., Computation, N. S.-J. of analysis and, & 2019, undefined. (2019). Heart disease prediction using data mining techniques. *Hal.Archives-Ouvertes.Fr*. https://hal.archives-ouvertes.fr/hal-02196156/

Ao, S.-I., Douglas, C., Grundfest, W. S., Burgstone, J., & International Association of Engineers. (2014). *World Congress on Engineering and Computer Science : WCECS 2014 : 22-24 October, 2014, San Francisco, USA*.

Aswale, S., & Shetgaonkar SRIEIT-Goa Shailendra Aswale, P. (2021). *Heart Disease Prediction using Data Mining Techniques*. www.ijert.org

Atallah, R. & Al-Mousa, A. (2019). Heart disease detection using machine learning majority voting ensemble method. *Ieeexplore.Ieee.Org*. Retrieved December 28, 2021, from https://ieeexplore.ieee.org/abstract/document/8923053/

Banu, G., Babikar, T., Bashier, I., of, N. S.-I. J., & 2019, undefined. (2019). Implementation of Artificial Neural Networks and Decision Tree Algorithms for Heart Disease Diagnosis. *Ijpras.Com*, *8*(2), 104–110. https://ijpras.com/storage/models/article/JDp6bQ3WqE8waQEdHXUMH4p4RmKNSzlN17gIuL823hM2eUsf54aCXxZQPFul/implementation-of-artificial-neural-networks-and-decision-tree-algorithms-for-heart-disease-diagno.pdf

Basha, N., PS, A., International, P. V.-2019 4th, & 2019, undefined. (2019). Early Detection of Heart Syndrome Using Machine Learning Technique. *Ieeexplore.Ieee.Org*. https://ieeexplore.ieee.org/abstract/document/9114651/

Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021).

Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning. *Computational Intelligence and Neuroscience*, *2021*. https://doi.org/10.1155/2021/8387680

Biology, R. S.-A. of the R. S. for C., & 2021, undefined. (2021). Proposed Random Forest Algorithm Using Heart Disease Prediction In Data Mining Process. *Annalsofrscb.Ro*, *12*(6), 4013–4023. https://annalsofrscb.ro/index.php/journal/article/view/9194

Borges, A. F. S., Laurindo, F. J. B., Spínola, M. M., Gonçalves, R. F., & Mattos, C. A. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, *57*, 102225. https://doi.org/10.1016/J.IJINFOMGT.2020.102225

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/J.IJINFOMGT.2021.102383

Damen, J. A. A. G., Hooft, L., Schuit, E., Debray, T. P. A., Collins, G. S., Tzoulaki, I., Lassale, C. M., Siontis, G. C. M., Chiocchia, V., Roberts, C., Schlüssel, M. M., Gerry, S., Black, J. A., Heus, P., Van Der Schouw, Y. T., Peelen, L. M., & Moons, K. G. M. (2016). Prediction models for cardiovascular disease risk in the general population: Systematic review. *BMJ*, *353*. https://doi.org/10.1136/bmj.i2416

David, H. B. F. & Belcy, S. A. (2018). HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES. *Ictactjournals.In*. Retrieved December 28, 2021, from http://ictactjournals.in/paper/IJSC_Vol_9_Iss_1_Paper_6_1817_1823.pdf

Decision Tree Regression — scikit-learn 1.0.2 documentation. (n.d.). Retrieved January 25, 2022, from https://scikit-learn.org/stable/auto_examples/tree/plot_tree_regression.html#sphx-glr-auto-examples-tree-plot-tree-regression-py

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., … Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994. https://doi.org/10.1016/J.IJINFOMGT.2019.08.002

G, M. (2020). Heart Disease Prediction Using Machine Learning Algorithms. *Bioscience Biotechnology Research Communications*, *13*(11), 24–27. https://doi.org/10.21786/bbrc/13.11/6

Gazeloğlu, C. (2020). Prediction of heart disease by classifying with feature selection and machine learning methods. *Progress in Nutrition*, *22*(2), 660–

670. https://doi.org/10.23751/pn.v22i2.9830

Hariharan, K., Vigneshwar, W. S., Sivaramakrishnan, N., & Subramaniyaswamy, V. (2018). A comparative study on heart disease analysis using classification techniques. *Hal.Archives-Ouvertes.Fr*. Retrieved December 28, 2021, from https://hal.archives-ouvertes.fr/hal-01826700/

Ibrahim, I., & Abdulazeez, A. (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *Journal of Applied Science and Technology Trends*, *2*(01), 10–19. https://doi.org/10.38094/jastt20179

Jindal, H., Agrawal, S., Khera, R., Jain, R & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. *Iopscience.Iop.Org*. https://doi.org/10.1088/1757-899X/1022/1/012072

Joshi, S., & Nair, M. K. (2015). Prediction of heart disease using classification based data mining techniques. *Smart Innovation, Systems and Technologies*, *32*, 503–511. https://doi.org/10.1007/978-81-322-2208-8_46

K-Means Clustering for Analysis of Heart Disease | by michelle hottinger | Medium. (n.d.). Retrieved January 25, 2022, from https://medium.com/@michellibelly/k-means-clustering-for-analysis-of-heart-disease-c2c6f75927e0

K-Nearest Neighbor(K-NN) | Python - AI ASPIRANT. (n.d.). Retrieved January 25, 2022, from https://aiaspirant.com/k-nearest-neighbor/

Kalamma, G., & Shreedhara, K. (2016). Predicting Heart attack using Data mining Technique. *Academia.Edu*. Retrieved December 28, 2021, from https://www.academia.edu/download/48538968/IJSARTV2I74622.pdf

Kavitha, M., Gnaneswar, G., Dinesh, R., Rohith Sai, Y & Sai Suraj, R. (2021). Heart Disease Prediction using Hybrid machine Learning Model. *Ieeexplore.Ieee.Org*. Retrieved December 28, 2021, from https://ieeexplore.ieee.org/abstract/document/9358597/

Kumar, N. K., Sindhu, G. S., Prashanthi, D. K., & Sulthana, A. S. (2020). Analysis and Prediction of Cardio Vascular Disease using Machine Learning Classifiers. *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, 15–21. https://doi.org/10.1109/ICACCS48705.2020.9074183

Padmanabhan, M., Yuan, P., … G. C.-J. of clinical, & 2019, undefined. (2019). Physician-friendly machine learning: A case study with cardiovascular disease risk prediction. *Mdpi.Com*, *8*(7). https://doi.org/10.3390/jcm8071050

Pandita, A., Vashisht, S., Tyagi, A., & Yadav, S. (2021). *Review Paper on Prediction of Heart Disease using Machine Learning Algorithms*. *9*. https://doi.org/10.22214/ijraset.2021.35626

Qaiser, H., Shafique, U., Majeed, F., Qaiser, H., & Mustafa, I. U. (2015). DATA MINING IN HEALTHCARE FOR HEART DISEASES. In *International Journal of Innovation and Applied Studies* (Vol. 10, Issue 4). http://www.ijias.issr-journals.org/

Random Forest Algorithm with Python and Scikit-Learn. (n.d.). Retrieved January 25, 2022, from https://stackabuse.com/random-forest-algorithm-with-python-and-scikit-learn/

Research, U. D.-B., & 2018, undefined. (2018). Prediction system for heart disease using Naive Bayes and particle swarm optimization. *Researchgate.Net*. https://doi.org/10.4066/biomedicalresearch.29-18-620

Sajja, G. S. (2021). A Comprehensive Review of Various Machine Learning Techniques for Heart Disease Prediction. *Article in International Journal of Computer Applications*, *183*(37), 975–8887. https://doi.org/10.5120/ijca2021921772

Shah, D., Patel, S., & Bharti, S. K. (2020). Heart Disease Prediction using Machine Learning Techniques. *SN Computer Science*, *1*(6), 345. https://doi.org/10.1007/s42979-020-00365-y

Smiti, A. (2020). When machine learning meets medical world: Current status and future challenges. *Computer Science Review*, *37*, 100280. https://doi.org/10.1016/J.COSREV.2020.100280

Yahaya, L., David Oye, N., & Joshua Garba, E. (2020). A Comprehensive Review on Heart Disease Prediction Using Data Mining and Machine Learning Techniques. *American Journal of Artificial Intelligence*, *4*(1), 20. https://doi.org/10.11648/j.ajai.20200401.12